

Figure 1

The novel gene as identified through RACE analysis (894 bp)

GGGAGTGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAA
AGCTTGCCGAACATAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAG
CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAT
GAAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCC
TGTCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGG
TGAAAATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCA
ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTCT
TCAGTTCCAACAAAAGGTCTGTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTG
AAGGAAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAT
GATGAGAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAAC
TGGAACACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATT
GCCTGATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTC
TTGGTCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGG
GAGCATGTACCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCA
GCTTTAAGGTTGTTGTGTTTTGTTTTGATTATGTTGCTTGTTAATAAAAAAAAAATAGAAA

A

Figure 2

Amino acid sequence as translated from the novel gene (210 amino acids)

MATETVELHKLKLAELKQECLARGLETGKGIKQDLIHRLQAYLEEHAEEEEANEEDVLGDETEEE
ETKPIELPVKEEEPPEKTVDVAAEKKVVKITSEIPQTERMQKRAERFNVVPSLESKKAARAAR
FGISSVPTKGLSSDNKPMVNLDKLERAQRFGLNVSSISRKSEDDEKLKKRKERFGIVTSSAG
TGTTEDEAKKRKRAERFGIA

Underlined sequences are amino acid sequences obtained by MS/MS analysis.

Figure 3

The sequence of the novel gene amplified through long distant PCR and used to construct the expression vector (873 bp).

TGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT
GCCGA ACTAAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAAGA
TCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAATGAAG
AAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCCTGTC
AAAGAGGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGCAGAGAAGAAAGTGGTGAA
AATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCAATGT
ACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAG
TTCCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGG
AAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATG
AGAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAACTGGA
ACCACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATTGCCT
GATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTCTTGG
TCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGGGAGC
ATGTACCCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCAGCTT
TAAGGTTGTTGTGTTTTTGTGTTTTGATTATGTTGCTTGTTAAT

Figure 4

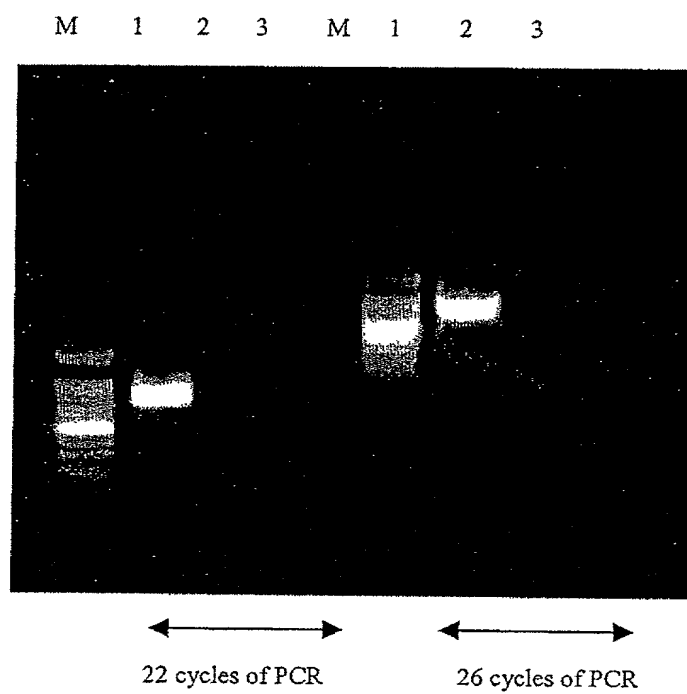


Figure 5

P-151 5'-Untranslated Region

```
1                                     75
CAGGGGCAGCAGTGATTATCTGAACTCGGATCTTTAAAATTGTGGTAGCTCTAAAGCTGATGATGTCTGGTTAGG
*****
76                                     150
AAGTGGCTCTTGCCCCGCCCCAGCCCCACCGCCAGTTCCTTAAGCCCGCCCCATGCCCCCTCCAGCTTCCTCCTCA
*****
151                                     225
TGTTTCATCGGTTTTTTTCAGGGCTCCCTTCAACGCTCCCCTCTCAGTATTTAGGTCACCACTCCCTCGGCGCCCCCT
*****
226                                     300
TTCGCCTCCCACCATTTTTCTCTCAGCAACCCTTACAGTCTTTGCAGCTCCTACCTGCCAGCTCAGATCCCCGTCC
*****
301                                     375
GGCTATGGGCGCGGCGCCGGCTACCACACCTGAAGTCTCCAGGAAGTAACGCCTCTCCTTCTGCCCCCTTTCCTGT

376                                     450
TGGAGGAACAGAATCAGCGCTGCCACCACCCATTGGTTGGTGGTCTGTAATGCAAGCACAGTTGGTTGCCATT

451                                     525
TCTGTCGTTGCAAGATACAGTGCCCCGCCCTCTCCAGTTCACCTTTTGAAAGAGGTGGGGCAAGCTGCCTAG

526                                     600
AGAAGTGAGAGCGACGTCAGCTATTGACCAATGGGAAGAGCTGATGGTATGGCGTGGGAGCAAGAGTGACAACGA

601                                     675
TTGGTCAGCCTTGCATCTCTACGCCTAAGGCGGGAACCTCTGGAGGCGGAGGCCGCGGGTGGGGGGAGTGGAGTG

676
AGGGGTAACAAGATG.....P151 coding region.....
```

(Total length: 690 bp)

Sequence with asterisk: the 274 bp fragment

Underlined sequences are the minicistrones or uORFs before the start of the P151 coding region with the start and stop codons in bold.